DOCUMENT RESUME

ED 232 580 HE 016 504

AUTHOR Bingman, Richard M.
TITLE Use of Student Evaluations to Improve Instruction. AIR 1983 Annual Forum Paper.
PUB DATE May 83
NOTE 17p.; Paper presented at the Annual Forum of the Association for Institutional Research (Toronto, Ontario, May 23-26, 1983).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Higher Education; *Inservice Education; Institutional Research; *Instructional Improvement; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; *Teaching Assistants; *Teaching Skills
IDENTIFIERS *AIR Forum

ABSTRACT

A pilot project designed to link more effectively instructional improvement to student evaluation was undertaken at The University of Missouri at Kansas City. Attention was focused on graduate teaching assistants (GTAs) in seven academic units. Students rated their GTA instructors, and interpretations of the evaluation results were provided to a sample of the GTAs. At the same time, the GTAs were provided followup assistance to help improve their instructional practices (i.e., seminars on how to improve lecturing, set objectives, utilize feedback, and design tests). A comparison group of GTAs received a printout of the data analysis, but no followup interpretation or training services. A pre-post (ex post facto) experimental design was utilized to compare student evaluations for 1 year for five experimental group GTAs and seven comparison group GTAs. The results indicate that the provision of interpretations of student ratings and followup assistance were effective in reaching six of seven objectives. Implications and future issues for colleges and universities using student reactions for general staff improvement, peer-tutorial, and supplemental programs are discussed. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED232580

USE OF STUDENT EVALUATIONS TO IMPROVE INSTRUCTION

BY
Richard M. Bingman, Ed.D
Program Associate in Higher Education
Mid-continent Regional Educational Laboratory
4709 Belleview
Kansas City, Missouri 64112
(816) 756-2401

Prepared for
Association for Institutional Research Forum
Toronto, Ontario
May 23-26, 1983

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

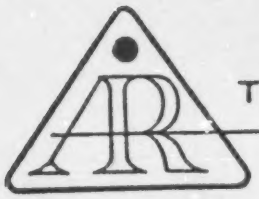
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

NE 016 504



THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the Twenty-Third Annual Forum of the Association for Institutional Research held at the Sheraton Centre in Toronto, Ontario, May 23-26, 1983. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum papers.

D. R. Coleman, Chairman
Forum Publication
Advisory Committee

USE OF STUDENT EVALUATIONS TO IMPROVE INSTRUCTION

Abstract

This study describes a pilot project designed to link more effectively instructional improvement to student evaluation. The pilot project on the campus of The University of Missouri at Kansas City focused on 25 graduate teaching assistants (GTAs) in seven academic units. A pre-post (ex post facto) experimental design was utilized to compare student evaluations for 12 GTAs within and outside the project over the period of one year. The results showed the treatment was effective for reaching six of seven objectives. Implications and future issues for colleges and universities using student reactions for general staff improvement, peer-tutorial and supplemental programs are discussed. A manual better linking evaluation and instructional improvement was produced and distributed to the units.

Background of the Problem

"Teachers learn about their classroom effectiveness both from what their students do and from what they say" (Kulik, 1976). In other words, one assessment of an instructor's performance is through testing student achievement. Another is through surveying student reactions to the instructor, which is the method reported here.

At The University of Missouri at Kansas City (UMKC), questionnaires on graduate teaching assistants' (GTAs) effectiveness had been administered regularly, but there was little or no apparent evidence that the results had been used in a systematic way to improve instruction. In a campus survey conducted in Spring, 1982, nearly 50 percent of the students who had completed questionnaires on teaching practices reported that they felt their evaluations had been of no benefit.

There are reservations about the value of student evaluations of instruction. According to Centra (Centra, 1981, Ch.2), these generally are:

- Students are not good judges of teaching.
- Student responses are unreliable.
- Teachers will not value student reactions sufficiently to make change.
- Student ratings are seriously contaminated by student/class characteristics.

In response to these points, some research findings are:

- "As a rule, student evaluations of classroom performance are more reliable than teacher peer evaluation" (Kerr, 1982).
- Seventy percent of faculty in a national survey agreed that their promotions should be based in part on student evaluations of their teaching (Cited in Centra, 1981, p.17).

- Student or class characteristics such as sex, grade point average, college year, age, and academic ability have weak or insignificant relationships to ratings of teaching effectiveness (Centra and Creech, 1976).
- A study of 400 instructors in five colleges reported that instructors showed improvement after half a semester if they had help interpreting their ratings (Centra, 1981).

Design of Project

Questionnaires traditionally had been administered to students in GTA-taught classes in the academic units of Arts and Sciences, Administration, and Education. The students were asked to rate their GTA instructor on seven items. The data were collected, analyzed and the results reported back to the unit chairpersons. The University provided no organized follow-up assistance.

During the 1981-82 academic year, the Division of Academic Affairs selected seven academic units with a total of 25 GTAs with teaching responsibilities to form a pilot treatment group. The criteria for selection were 1) units which ranked low in evaluations and 2) units which represented a cross-section of academic disciplines.

The treatment consisted of providing 1) the GTAs with interpretations of their evaluation results and 2) follow-up assistance to improve their instructional practices. The assistance consisted of a general seminar on "How to Improve Lecturing," and departmental seminars on setting objectives, questioning strategies, receiving and utilizing feedback, and designing tests. The remaining non-treatment academic units, which had 45 GTAs with teaching responsibilities, did receive a print-out of the data analysis, but did not receive any follow-up interpretation or training services. The investigators

conducted a telephone survey of the non-treatment units and determined that no use was made of the evaluation results beyond an occasional review by the chairperson.

Design of Study

For this study, data for comparisons of treatment versus non-treatment groups were based on the Spring, 1981, and Spring, 1982, student evaluations. Because of the rapid turnover of GTAs and incomplete data collections, only 12 were evaluated both times, five in the treatment and seven in the non-treatment group. The two groups were compared on seven items contained in the questionnaire. Five of the items, labeled "Core Group," were considered as assessments of direct outcomes of the project; two, labeled "Peripheral Group," were considered less directly related, but represented desired outcomes.

TABLE 1

CORE ITEMS

4. The instructor clearly communicates directions, instructions and assignments.
5. The instructor is careful and precise in explaining material and answering questions.
6. The instructor engages in and leads class discussion effectively.
8. The instructor's dedication to high standards of teaching has had a stimulating effect upon my appreciation of the subject.

*9-10. I would recommend this instructor to other students.

* Item 9 was on the 1981 questionnaire; item 10 on the 1982 questionnaire.

TABLE 2

PERIPHERAL ITEMS

3. The instructor's language proficiency is such that he/she communicates clearly; I can understand what he/she says.

7. The instructor speaks and listens to students in a courteous and respectful manner.

Analysis of Data

Five responses were possible for each item: (1) strongly agree, (2) agree, (3) undecided, (4) disagree, (5) strongly disagree. These were re-categorized for analysis into two groups: (A) agreement - (1), (2) and (B) non-agreement - (3), (4) and (5).

These two sets of responses were divided into four groups:

(1) Treatment Group - Pre (N=183)

(2) Treatment Group - Post (N=170)

(3) Non-treatment Group - Pre (N=229)

(4) Non-treatment Group - Post (N=203)

The proportion of (A) agreement responses to total responses for each item in each group is calculated and expressed as a decimal rounded to three places.

The level of significance was determined by the Lawshe-Baker method, using a pre-determined level of confidence of $p<.05$, meaning that differences in proportions are likely to occur by chance fewer than five times in 100. (Downie and Heath, 1980).

Results

The findings were organized in terms of four questions:

1. Were there significant differences between the treatment and non-treatment groups on the pre-test data on any of the seven items?
2. Were there significant differences between the treatment and non-treatment groups on the post-test data on any of the seven items?
3. Were there significant differences in the pre-post data for the treatment group on any of the seven items?
4. Were there significant differences in the pre-post data for the non-treatment group on any of the seven items?

The following four Tables (Tables 3,4,5,6) present data in answer to the above four questions.

TABLE 3

PRE-TEST COMPARISONS

OF

TREATMENT AND NON-TREATMENT GROUPS

ITEM	TREATMENT (PRE) (N=183)	NON-TREATMENT (PRE) (N=229)	LEVEL OF SIGNIFICANCE
3	.804	.840	p >.05
4	.833	.856	p >.05
5	.825	.877	p >.05
6	.798	.817	p >.05
7	.825	.918	p <.05
8	.683	.807	p <.05
9	.730	.799	p >.05

There were no significant differences between the two groups at the start of the project on items 3, 4, 5, 6, and 9. There were significant differences in favor of the non-treatment group on items 7 and 8. Item 8 is a core item, "The instructor's dedication to high standards of teaching has had a stimulating effect upon my appreciation of the subject." Item 7 is a peripheral item, "The instructor's language proficiency is such that he/she communicates clearly: I can understand what he/she says." These results suggest that there were no extensive differences in the two groups at the beginning of the study.

TABLE 4

POST-TEST COMPARISONS

OF

TREATMENT AND NON-TREATMENT GROUPS

ITEM	TREATMENT (POST) (N=170)	NON-TREATMENT (POST) (N=203)	LEVEL OF SIGNIFICANCE
3	.960	.813	$p < .05$
4	.942	.793	$p < .05$
5	.923	.803	$p < .05$
6	.895	.759	$p < .05$
7	.959	.795	$p < .05$
8	.861	.700	$p < .05$
9	.941	.818	$p < .05$

There were significantly higher differences which favored the treatment group at the end of the study on all items. The largest gains were made on items 7 and 8 which favored the non-treatment group at the beginning of the study (see Table 3). These results suggest that improved communication among GTAs and students may be a key factor towards improving the students' perceptions and attitudes towards GTA practices and the subject itself.

TABLE 5

COMPARISON OF

PRE- AND POST-TREATMENT GROUPS

ITEM	TREATMENT (PRE) (N=183)	TREATMENT (POST) (N=173)	LEVEL OF SIGNIFICANCE
3	.804	.960	$p < .05$
4	.833	.942	$p < .05$
5	.825	.923	$p < .05$
6	.798	.895	$p > .05$
7	.825	.959	$p < .05$
8	.683	.861	$p < .05$
9	.730	.941	$p < .05$

There were significantly higher differences in the pre-post data for the treatment group on six of seven items. The one other difference was higher but insignificant. This was core item No. 6 in the student questionnaire, "The instructor engages in and leads classroom discussion effectively." A further check on the breakdown of the undesired responses showed that a high percentage of students were "undecided" on this item. This suggests 1) the GTA behaviors remained unchanged, 2) the students did not recognize the changes in behavior within an instructional mode which continued to favor lecturing, and 3) the interpretation of such a broad concept varies tremendously among the students. These findings suggest that this GTA training objective be given more attention in terms of improved clarity and more congruence between objectives, instructional strategies and assessment.

TABLE 6

COMPARISON OF

PRE- AND POST-NON-TREATMENT GROUPS

ITEM	NON-TREATMENT (PRE) (N=229)	NON-TREATMENT (POST) (N=203)	LEVEL OF SIGNIFICANCE
3	.840	.813	$p > .05$
4	.856	.793	$p > .05$
5	.877	.803	$p > .05$
6	.817	.759	$p > .05$
7	.918	.795	$p < .05$
8	.807	.700	$p < .05$
9	.799	.818	$p > .05$

There were no significant differences in the pre-post data for the non-treatment group on five of seven items. A significantly negative change was reported on peripheral item No. 7, "The instructor speaks and listens to students in a courteous and respectful manner," and No. 8, "The instructor's dedication to high standards of teaching has had a stimulating effect upon my appreciation of the subject." This finding is particularly interesting since the non-treatment group was significantly higher than the treatment group on these items at the beginning of the study. These results combined with those presented in Table 4 suggest that the training program provided additional impact in areas which would decline if left alone.

Discussion

The findings on Tables 3 and 4, the pre- and post-test comparisons, suggest that GTAs do not improve their instructional effectiveness merely through having additional teaching experience, as GTAs receiving follow-up assistance had better student ratings than those left alone. The results on Table 6, a comparison of the pre- and post-test non-treatment groups, show that the non-treatment GTAs even received somewhat lower ratings on the second evaluation than on the first.

These findings have important implications for UMKC and other institutions of higher education concerned about the quality of their teaching. At UMKC a large percent of freshman classes are assigned to GTAs having little or no teaching preparation. Since the teachers of freshman classes are often the students' most important contact with the university, the quality of their instruction is vital to both students and the institution. This factor becomes more crucial when many of the classes assigned to GTAs are labeled as "high risk," meaning that many students are likely to drop out. The results of this study indicate that students in the treatment group improve their attitudes toward the GTA and the subject whereas without this support the attitudes of students towards the GTA and the subject may even decline.

The findings on Table 5 show that only the class discussion strategy failed to reach the hoped-for significance level. This finding is important for classes in which lecturing is the only accepted mode. Since concept formation is usually regarded as important in higher education, alternative instructional deliveries might be considered, such as small groups, peer teaching and micro-computer supplements. Better models are needed to mix or blend these strategies and study their effects with students demonstrating wide ranges of

interests and abilities.

The results of this study point out the importance of improved communication throughout an instructional program based on student-oriented goals and objective strategies for attaining them. A report of the college outcomes measures project (Forrest, 1982), points out that institutions having both features are intent on making it clear to all concerned what difference general education should make in students. As a result, faculty and students are provided important organizers which give direction to teaching and learning efforts. Further, Forrest concludes it is probable that there is a relationship between these factors and increased student learning.

To facilitate communication at UMKC concerning student-oriented goals and objectives, a procedures manual was developed to effectively link data collection and utilization procedures.* It has been distributed to all academic units in the Arts and Sciences Division and efforts are under way to study its utilization. This is the first step in a series of planned actions to effectively link modern techology with human resources. The ultimate outcome is to improve teaching and learning in higher education.

* Copy available from Terry Cezar, 100 Student Academic Support Service Building, University of Missouri at Kansas City, 5100 Rockhill Road, Kansas City, Missouri 64110.

BIBLIOGRAPHY

Addy, C. Some Thoughts on Faculty Evaluation. Innovation Abstracts, December 4, 1981, 3-31.

Bergquist, W.H., Phillips, S.R. A Handbook for Faculty Development. Vol. 1, Washington D.C.: Council for the Advancement of Small Colleges, 1977. Vol. 2, Berkeley, Calif.: Pacific Soundings Press, 1977.

Bingman, R.M., Cezar, T., and Chamberlin, M.E. Procedures Manual: Graduate Teaching Assistants Evaluation, Administration, and Utilization. Kansas City, Mo.: Learning Center, University of Missouri at Kansas City, 1982.

Centra, J.A., Determining Faculty Effectiveness. San Francisco: Jossey-Bass, 1981.

Centra, J.A., and Creech, F.R. The Relationship Between Student, Teachers, and Course Characteristics and Student Ratings of Teacher Effectiveness. Princeton, N.J.: Project Report 76-1, Educational Testing Service, 1976.

Centra, J.A. The Utility of Student Ratings for Instructional Improvement. Princeton, N.J.: Project Report 72-16, Educational Testing Service, 1972.

Downie, N.M., and Heath, R.W. Basic Statistical Methods. New York: Harper and Row, 1980.

Eble, K.E. The Craft of Teaching. San Francisco: Jossey-Bass, 1976.

Forrest, A. Increasing Student Competence and Persistence: The Best Case for General Education. Iowa City, Ia.: National Center for the Advancement of Educational Practices, American College Testing, 1982.

How to Succeed as a New Teacher. New Rochelle, N.Y.: Change Magazine Press, 1978.

Hansen, S., and Hansen, W. Teacher Training Programs in College Economics: Their Development, Current Status, and Future Prospects. Journal of Economic Education, 1980, 11-2, 150-162.

Highet, G. The Art of Teaching. New York: Random House, 1950

Kerr, D. Student Evaluations of Teaching: Uses and Limitations. Innovation Abstracts, March 12, 1982, 4-8.

Kulik, J.A., and McKeachie, W.J. The Evaluation of Teachers in Higher Education. In F.N. Kerlinger (Ed.). Review of Research in Higher Education. New York: Peacock, 1975.

Kulik, J.A. Student Reactions to Instruction. Memo to the Faculty, October, 1976, 5-8.

McKeachie, W.J. Teaching Tips: A Guidebook for the Beginning College Teacher. Lexington, Mass.: D.C. Heath, 1977.

Rouche, John E. Holistic Literacy in College Teaching. New York: Media Systems Corporation, 1980.